

# 国立国語研究所学術情報リポジトリ

## 『中納言』の概要

メタデータ	言語: jpn 出版者: 公開日: 2021-03-09 キーワード (Ja): キーワード (En): 作成者: 大村, 舞, 石本, 祐一, 岡, 照晃 メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00003206">https://doi.org/10.15084/00003206</a>

# 『中納言』の概要

大村舞・石本祐一・岡照晃 (国立国語研究所 コーパス開発センター)

## コーパス検索システム『中納言』

中納言はコーパスを単語や品詞などの形態論情報で検索可能な検索システム (要ユーザ登録)



### 9つのコーパスについて検索可能

### 検索結果の例

検索語の品詞情報、検索語が現れる資料の情報を表示

コーパス名	種別	個別検索	包括検索	備考
現代日本語書き言葉均衡コーパス 中納言版	BCCWJ	●	●	従来より利用いただいていたBCCWJのデータです(コーパスの紹介ページ)。こちらのページからBCCWJ/アンダーソンデータをダウンロードできます。笑天版はこちら
国語研日本語ウェブコーパス 中納言版	NWJC	●	●	笑天版はこちら
日本語話し言葉コーパス	CSI	●	●	コーパスの紹介ページ
日本語日常会話コーパス モニター公開版	CEJC	●	●	コーパスの紹介ページ
昭和話し言葉コーパス モニター公開版	SSC	●	●	コーパスの紹介ページ
老大会話コーパス	NUCC	●	●	コーパスの紹介ページ
東日研・機構談話コーパス	CWPC	●	●	コーパスの紹介ページ
日本語歴史コーパス	CHU	●	●	コーパスの紹介ページ
日本語方言コーパス モニター公開版	COJADS	●	●	コーパスの紹介ページ
多言語母語の日本語学習者標本コーパス	I-JAS	●	●	I-JASに関する詳細な情報は「I-JAS 連携ページ」をご参照ください。ファイルテキスト・音声・ビデオはデータの読み込みからダウンロードできます。

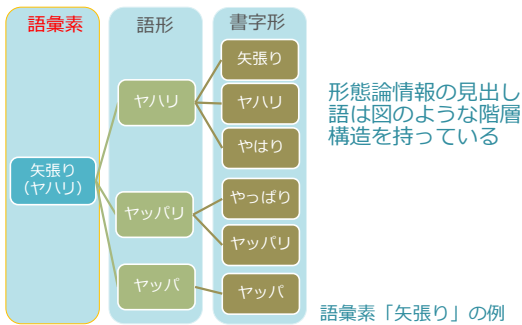
38,490件の検索結果が見つけられました。そのうち500件を表示しています。表示している(新)NDCの由来:国立国語研究所ウェブサイト(https://iss.nijl.go.jp/information/napi/2019年4月30日現在)に基づき、『現代日本語書き言葉均衡コーパス』(以下、種別)サンプルのNDCを再確認した。

検索対象語: 124,100,954  
 記号: 構文記号: 空白記号: 104,911,450

サンプルID	開始位置	漢字	平仮名	後文脈	品詞	活用型	活用形	レキシター	書名	著者	出版者	出版年
LBK2_00050	13140	7740	ju	*あの対象が、15インチは、こば	名詞	普通名詞一般		ローカル・クラスター	父の罪	ローカル・クラスター	文芸春秋	1966
LBH7_00016	7800	4720	*あの目標は、社会生活そのものが	名詞	普通名詞一般		ローカル・クラスター	佐々木 基一(著)	検査の要務	佐々木 基一(著)	講談社	1993
PNAL_00012	1750	1110	て彼が夢見ると、夢が実現でき	名詞	普通名詞一般		出版	伏原 的知子(著) 北山 千秋(著)	京野新聞	京野新聞社	2004	

### 形態論情報

語彙素という単語の見出し語に付与されている情報



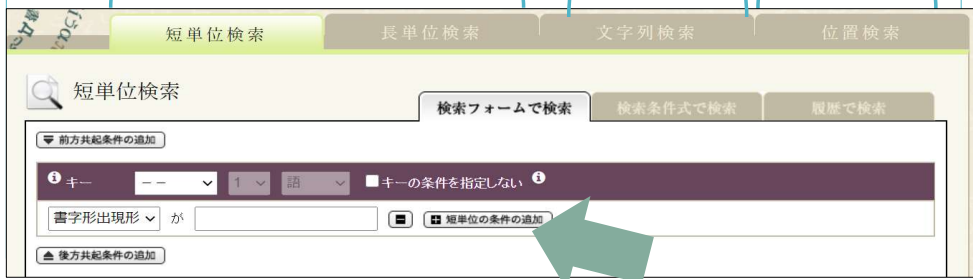
語彙素: 辞書の見出しのレベル  
 語形: 異語形を区別するレベル  
 書字形: 異表記を区別するレベル

語彙素を使うことで表記の違いをまとめることができる

### 中納言の4つの検索方法

中納言では4種類の検索方法によって検索可能 (一部コーパスでは長単位検索はない)

短単位検索・長単位検索: 形態論情報を使った詳細検索  
 文字列検索: 形態論情報を使わず出現形で検索  
 位置検索: 出現位置から (再) 検索



細かく情報を指定して検索可

### 短単位検索・長単位検索

形態論情報に基づいた検索ができる

#### 短単位検索

短単位が一致する検索が可能



語彙素「日本」の検索例

語彙素「日本」がヒットする → 「日本」「ニッポン」「にっぽん」「にほん」いずれも共通語彙素なのでヒット

#### 長単位検索

長単位が一致する検索が可能

「日本語」「日本語研究」「日本語教育」などが長単位



### 文字列検索

単語単位を気にせず完全一致する検索が可能



表記の違い (読みや書き方の違い) は考慮しないため その場合は短単位・長単位検索を利用する必要がある

### 位置検索



「ID」と「開始位置」でコーパス中の用例が一意的に決まるため、2つを保持していれば再現検索可能